

SPARSE SIGNAL AND IMAGE RECOVERY FROM COMPRESSIVE SAMPLES

Emmanuel Candès, Nathaniel Braun, Michael Wakin

Applied and Computational Mathematics, California Institute of Technology

ABSTRACT

In this paper we present an introduction to Compressive Sampling (CS), an emerging model-based framework for data acquisition and signal recovery based on the premise that a signal having a sparse representation in one basis can be reconstructed from a small number of measurements collected in a second basis that is incoherent with the first. Interestingly, a random noise-like basis will suffice for the measurement process. We will overview the basic CS theory, discuss efficient methods for signal reconstruction, and highlight applications in medical imaging.

Index Terms— Compressive sampling, sparse signal representations, random measurements, tomography, MRI

1. INTRODUCTION

Many natural phenomena give rise to signals having a sparse representation in some basis or dictionary. Consider, for example, a basis $\Psi = [\psi_1 \ \psi_2 \ \cdots \ \psi_N]$ for signals in \mathbb{R}^N ; two common possibilities are the Fourier and wavelet transforms. A digital signal $x \in \mathbb{R}^N$ can be represented as a linear combination $x = \Psi\alpha$ of the basis elements ψ_n (the columns of Ψ), with $\alpha \in \mathbb{R}^N$ providing the appropriate weights. Assuming a suitable choice of the basis Ψ , it is often the case in practice that many entries in α are approximately zero, with only $K \ll N$ coefficients carrying the significant information. Naturally, the conciseness of such a representation permits efficient signal processing and data compression. A more surprising and far-reaching implication, however, is that the assumed compressibility of the signal has significant bearings on the acquisition process itself. In fact, such signals can actually be acquired using far fewer samples than the signal size apparently demands.

In this paper, we present an introduction to Compressive Sampling (CS) [1–3], an emerging model-based framework for data acquisition and signal recovery in which sparse signals can be reconstructed from very small numbers of measurements — far fewer than the signal size N and proportional instead to the sparsity level K . A key requirement of the CS samples is that they be collected in a measurement

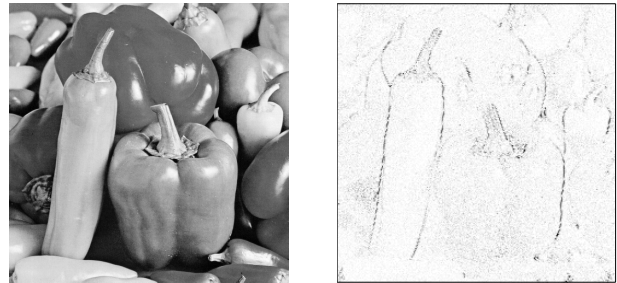


Fig. 1. Peppers image and one subband of its wavelet transform (magnitudes shown; vertical subband). The relatively few significant wavelet coefficients tend to cluster around the edges of the objects in the image. Many such images will be highly compressible, with the locations of significant coefficients changing from image to image.

basis Φ that is incoherent with the sparsity basis Ψ . Interestingly, a random noise-like basis Φ will suffice for the measurement process, as will a partial Fourier transform for MR imaging applications.

2. COMPRESSIVE SAMPLING (CS)

2.1. Sparse and compressible signal models

Many natural signals have concise representations when expressed in the proper basis. Consider, for example, the *Peppers* image in Figure 1 and its wavelet transform. Although nearly all pixels in the original image have nonzero values, the wavelet coefficients offer a concise summary: most wavelet coefficients are small, and the relatively few large coefficients (clustered around edges in the image) are all that is required to reconstruct the original image pixels with high accuracy.

Summarizing the above mathematically, consider a signal $x \in \mathbb{R}^N$ (such as an N -pixel image) and let $\Psi = [\psi_1 \ \psi_2 \ \cdots \ \psi_N]$ be an orthonormal basis (such as a wavelet basis) for signals in \mathbb{R}^N . We can write $x = \Psi\alpha$, where α is an $N \times 1$ coefficient vector that can be computed from the signal x :

$$\alpha_n = \langle \psi_n, x \rangle, \quad n = 1, 2, \dots, N. \quad (1)$$

CS relies on a specific model for the signal x , namely that it be sparse or compressible in the basis Ψ . By *sparse* we mean that the coefficient vector α has precisely K nonzero entries, where potentially $K \ll N$. (We write $\|\alpha\|_0 = K$; the ℓ_0 “norm” of a vector counts its nonzero entries.) By *compressible* we mean that the sorted magnitudes of the coeffi-

This work has been partially supported by National Science Foundation (NSF) grants ITR ACI-0204932 and CCF515362, NSF fellowship DMS-0603606, and the 2006 NSF Waterman Award.

cients in α decay quickly, and that α is well-approximated as an $N \times 1$ vector with just K nonzero entries.

Under either of these assumptions, the signal x has a very compact representation in terms of the significant coefficients in α ; a simple method for data compression would simply be to compute α from x (via (1), for example) and then encode the locations and values of the K significant coefficients. Indeed, this basic idea forms the core of modern signal and image compression schemes such as JPEG-2000 [4].

Unfortunately, this compression process requires (a) direct access to the sparse basis Ψ , and (b) computing *all* N of the coefficients α , as the locations of the significant coefficients may not be known in advance. The recent theory of CS suggests however, that the relevant information in the signal x can be captured using a small number of nonadaptive (even random) measurements of the signal. This suggests the potential for very efficient acquisition of sparse data, or equivalently, highly accurate recovery of sparse data from undersampled measurements.

2.2. Incoherent sampling

We suppose that, rather than computing the N coefficients α directly using (1), we collect $M < N$ measurements

$$y_m = \langle \phi_m, x \rangle, \quad m = 1, 2, \dots, M, \quad (2)$$

using M measurement vectors ϕ_m that are fixed in advance and not dependent on α (or more generally, on Ψ at all). We let Φ denote the $M \times N$ measurement matrix whose rows are the vectors ϕ_m^T ; we have then that the compressive measurements $y = \Phi x$, where $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$.

Under certain conditions on the number of measurements M and the matrix Φ , it is possible to recover the signal x from the compressive measurements y , even though (a) the length of y may be significantly smaller than x , and (b) the measurement matrix Φ may look nothing like the sparse basis Ψ . Indeed, because we do not permit Φ to depend on α , this second property is actually critical — in the *Peppers* image, for example, if Φ simply contained M arbitrary wavelet basis functions (where $M \ll N$), many significant wavelet coefficients would likely go completely unmeasured. Instead it is important that each measurement vector ϕ_m combine information from all of the sparse vectors ψ_n . There are several ways of quantifying this notion of incoherence. The essential requirement, when Φ is chosen as a random $M \times N$ submatrix of an $N \times N$ measurement basis, is that the rows of the measurement basis and the columns of the sparsity basis have small inner product; we refer the reader to [5] for additional details and discussion.

Suitable incoherence is assured for various useful pairs (Ψ, Φ) , for example when Ψ is the identity matrix (or the finer scales of a wavelet transform) and Φ is a random subset of $M = O(K \log N)$ rows from the discrete Fourier matrix. (This is essentially the conventional MRI setting [1, 6].) Interestingly, however, a measurement matrix Φ generated ran-

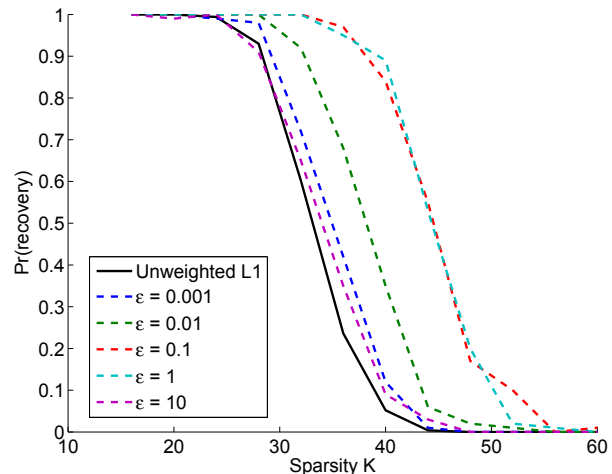


Fig. 2. Sparse signal recovery from $M = 100$ random measurements of a length $N = 256$ signal. The probability of successful recovery depends on the sparsity level K ; the dashed curves represent a reweighted ℓ_1 algorithm that outperforms the traditional unweighted ℓ_1 approach (solid curve).

domly with iid Gaussian or Bernoulli (± 1) entries will, with exceptionally high probability (assuming $M = O(K \log N)$), be incoherent with an arbitrary sparse basis Ψ .

2.3. Model-based recovery via ℓ_1 minimization

The process of recovering the signal x from the measurements y is ill-posed in general when $M < N$: there exist an infinite number of candidate signals \hat{x} for which $\Phi \hat{x} = \Phi x = y$. The key to CS recovery is to select among these candidates by imposing the *model* of sparsity or compressibility.

Supposing x is K -sparse in the dictionary Ψ , and assuming suitable incoherence between Ψ and Φ (and hence that $M = O(K \log N)$), then the convex optimization problem¹

$$\min \|\hat{\alpha}\|_1 \quad \left(= \sum |\hat{\alpha}_n| \right) \quad \text{such that } y = \Phi \Psi \hat{\alpha} \quad (3)$$

will return the correct coefficient vector $\hat{\alpha} = \alpha$ (and hence the correct signal $\hat{x} = \Psi \hat{\alpha} = x$). Figure 2 shows a simple experiment with $N = 256$. For several values of K , we construct K -sparse signals randomly in the time domain (Ψ is the identity matrix), assigning zero-mean, unit-variance Gaussian coefficient values to K nonzero elements in random positions. For each signal we set $M = 100$ and construct a random $M \times N$ measurement matrix Φ with iid Gaussian entries. The solid black curve denotes the probability (over 100 trials with random x and Φ) of correctly recovering the K -sparse signal from the measurements $y = \Phi x$. We see that, using $M = 100$ measurements, we can correctly recover K -sparse signals with very high probability when $K \approx 25 = M/4$.

It is important to stress that the CS generalizes well to more realistic settings. In cases where x is not strictly sparse

¹Problem (3) can be recast as a linear program and solved efficiently. See www.l1-magic.org for a complete software suite.

but rather compressible, the estimate $\hat{\alpha}$ obtained from (3) will approximate α with quality comparable to the best possible K -sparse approximation to α [2, 3]. Also, in the case of noisy measurements $y = \Phi x + z$, the ℓ_1 recovery program can be modified as follows

$$\min \|\hat{\alpha}\|_1 \text{ such that } \|y - \Phi \Psi \hat{\alpha}\|_2 \leq \epsilon, \quad (4)$$

yielding a reconstruction with provable accuracy bounds [7].

2.4. Impact and applications

CS can permit radically more efficient methods for data acquisition, as it suggests that data obeying a sparse or compressible model need only be sampled (via the measurement process Φ) at a rate proportional to its information level $K \ll N$. Indeed the potential randomness of Φ is particularly intriguing for applications in which the sparse basis Ψ is unknown at the encoder or impractical to implement for data compression — the knowledge and ability to implement Ψ are required only for the decoding or recovery of x . Viewed from a medical imaging perspective, the only way to acquire x may be to use a Φ of a certain modality (such as a subsampled Fourier transform); assuming a sparse basis Ψ exists for x that is also incoherent with Φ , efficient and accurate signal recovery will be possible. Still other implications of the CS theory concern error correction in communications channels [8] and distributed source coding for sensor networks [9].

3. ALTERNATIVE RECONSTRUCTION METHODS

As discussed above, the key to identifying the proper signal x from among an infinite number of possibilities is to use a model. The requirement of a small ℓ_1 -norm (as a measure of sparsity or compressibility) is one such model; it is generally effective, but also general purpose, and certainly not representative of the wide range of models developed over the years for signal and image processing. Let us briefly mention two possibilities for incorporating such models.

3.1. TV minimization

One alternative to ℓ_1 minimization is motivated not from the perspective of sparsity but from the (related) observation that many natural images have limited spatial variability. This is captured by the image's "total variation" (TV), the sum of the magnitudes of the discrete image gradient at each point:

$$\text{TV}(x) = \sum_{i,j} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}.$$

For an early use of the TV norm in image processing, see [10].

As an alternative to (3), the optimization problem

$$\min \text{TV}(x) \text{ such that } y = \Phi x \quad (5)$$

can be solved efficiently as a second-order cone program.

As an example, we consider the angiogram shown in the left panel of Figure 3. From this $N = 256 \times 256 = 65536$ pixel real-valued image x we collect $M = 9368$ complex-valued measurements y along 80 radial lines in the Fourier domain. In the center panel we show the result of simple backprojection. In the right panel we show the result of TV minimization, which produces a reconstruction with far fewer artifacts. For this experiment, we have supplemented the TV minimization with a nominal amount of side information about the ℓ_1 norm of the image's wavelet coefficients; for the full details of this experiment we refer the reader to [11].

3.2. Reweighted ℓ_1 minimization

Returning again to the setting of ℓ_1 minimization, we discuss a slight reformulation that improves the recovery of sparse signals. Historically, the motivation for choosing the ℓ_1 norm in (3) comes from the fact that it provides a convex relaxation of a more desirable but intractable optimization problem

$$\min \|\hat{\alpha}\|_0 \text{ such that } y = \Phi \Psi \hat{\alpha}. \quad (6)$$

Let W denote an $N \times N$ matrix with nonzero entries w_1, w_2, \dots, w_N along the diagonal and zeros elsewhere. We note that in (6), the norm $\|\hat{\alpha}\|_0$ can be replaced by the reweighted norm $\|W\hat{\alpha}\|_0$ without changing the solution. However, minimizing the corresponding relaxation

$$\min \|W\hat{\alpha}\|_1 \text{ such that } y = \Phi \Psi \hat{\alpha}, \quad (7)$$

will not return the same solution as the unweighted norm $\|\hat{\alpha}\|_1$ in (3).

Our goal is to obtain a set of weights W that actually improve upon the ℓ_1 norm, making it behave more like the ℓ_0 norm. The key difference between the ℓ_1 and ℓ_0 norms is the dependence on magnitude — larger coefficients are penalized more heavily than smaller coefficients, unlike the democratic penalization of the ℓ_0 norm. Ideally the weights W could correct for this imbalance, by setting w_n inversely proportional to α_n . Of course, the true coefficients α are not known in advance, but an iterative procedure can be used that alternates between estimating $\hat{\alpha}$ and redefining the weights. The algorithm is as follows

1. Set the iteration $i = 1$. Set $w_n^{(1)} = 1$ for $n = 1, 2, \dots, N$.
2. Let $\hat{\alpha}^{(i)}$ be the solution to (7) with weights $w_n = w_n^{(i)}$.
3. Set $w_n^{(i+1)} = (|\hat{\alpha}_n^{(i)}| + \epsilon)^{-1}$.
4. Increment i and go to step 2.

The parameter ϵ in step 3 should be set slightly smaller than the expected entries of α . As demonstrated below, the recovery process is somewhat robust to the choice of ϵ .

As an alternative interpretation of this reweighted ℓ_1 recovery algorithm, it is interesting to note that every iteration of the above algorithm is guaranteed [12] to decrease the following function of $\hat{\alpha}$: $\sum_{n=1}^N \log(|\hat{\alpha}_n| + \epsilon)$. This objective

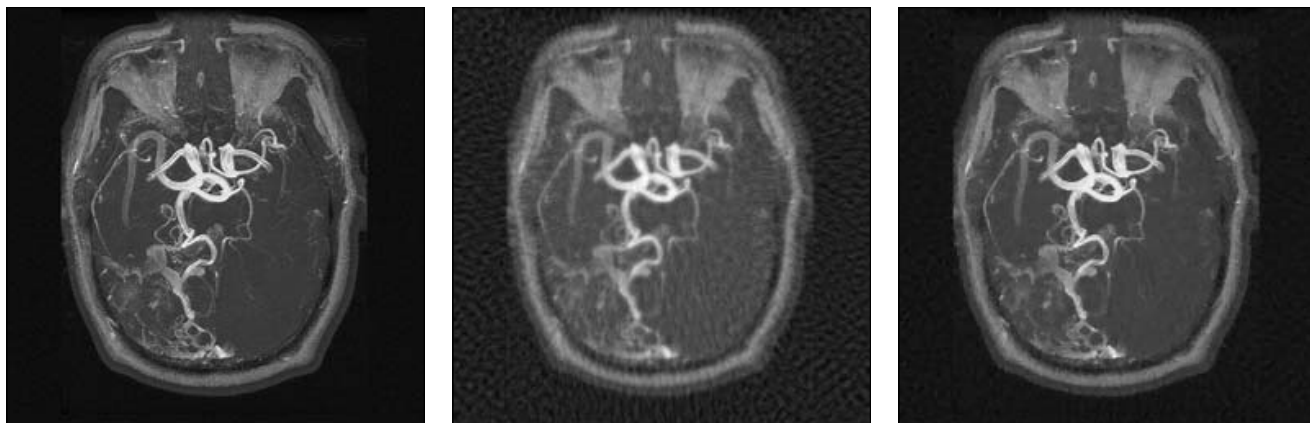


Fig. 3. Angiogram recovery from compressive (simulated MR) measurements. Left panel: original image. Center panel: backprojection from 80 lines in Fourier domain, $\text{PSNR} = 10 \log_{10}(255^2/\text{MSE}) = 22.15\text{dB}$. Right panel: TV minimization reconstruction, $\text{PSNR} = 26.81\text{dB}$.

function is concave (not convex) but more closely resembles the ℓ_0 norm. Unfortunately the algorithm is guaranteed only to converge to a local (not necessarily global) minimum.

Figure 2 shows the results of reweighting ℓ_1 for the experiment described in Section 2.3; the dashed curves represent several different values for the parameter ϵ . We see a marked improvement over the unweighted ℓ_1 algorithm — with the proper choice of ϵ , the ability to recover sparse signals has increased from $K \approx 25 = M/4$ to $K \approx 33 = M/3$.

Preliminary and ongoing work has also shown promising results from reweighted versions of other CS recovery algorithms. For example, in the case of noisy measurements we have experimented with a reweighted version of the quadratically constrained ℓ_1 program (4). For moderate noise levels, reweighting typically yields signal estimates with 30-40% lower mean-square error (MSE) than the unweighted algorithm. Additional details will appear in a subsequent paper.

4. CONCLUSIONS AND EXTENSIONS

CS provides a powerful model-based framework for recovering data from incomplete information and has many promising implications in data acquisition, medical imaging, remote sensing, etc. There is a vast array of literature available in this field, ranging from the practical to the mathematical, which we have been unable to properly survey in this short paper. We refer the reader to [13] for a more detailed survey of the key CS principles, to [14] for various possible extensions of the CS methodology, and to dsp.rice.edu/cs for a more comprehensive collection of CS references.

Our presentation has focused on sparse and compressible models, but as we mentioned in Section 3, similar concise models for signal structure can also be incorporated. Another such model, which we have omitted for brevity, concerns the case where the signal x depends on some K -dimensional parameter θ ; example parameters include the orientation and position of an edge within an image, the position of a camera photographing a scene, etc. In these cases, the set of possible

signals x traces out a K -dimensional manifold in the ambient signal space \mathbb{R}^N . For reasons highly related to the preceding discussion of incoherence, in certain situations manifold-modeled signals can be recovered from small numbers of compressive measurements. We refer the reader to [15, Chapter 6] for more detailed information and examples, including experiments aimed at improving image reconstruction from compressive measurements.

5. REFERENCES

- [1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, Feb. 2006.
- [2] E. Candès and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, Dec. 2006.
- [3] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, Apr. 2006.
- [4] D. S. Taubman and M. W. Marcellin, "JPEG 2000: Image compression fundamentals, standards and practice," *Kluwer*, 2001.
- [5] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," 2006, preprint.
- [6] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," 2007, preprint.
- [7] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, 2006.
- [8] E. J. Candès, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *Proc. IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005.
- [9] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," 2005, preprint.
- [10] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," vol. 60, pp. 259–268, 1992.
- [11] E. Candès and J. Romberg, "Signal recovery from random projections," in *Proc. SPIE Computational Imaging III*, 2005, vol. 5674, pp. 76–86.
- [12] M. Fazel, H. Hindi, , and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. American Control Conference*, 2003.
- [13] E. Candès, "Compressive sampling," in *Proc. International Congress of Mathematics*, Madrid, Spain, 2006.
- [14] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, Mar. 2006.
- [15] M. B. Wakin, *The Geometry of Low-Dimensional Signal Models*, Ph.D. thesis, Rice University, Houston, TX, 2006.